

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
8 November 2001 (08.11.2001)

PCT

(10) International Publication Number  
WO 01/84539 A1(51) International Patent Classification<sup>7</sup>: G10L 15/26Eindhoven (NL). STEINBISS, Volker; Prof. Holstlaan 6,  
NL-5656 AA Eindhoven (NL).

(21) International Application Number: PCT/EP01/04714

(22) International Filing Date: 26 April 2001 (26.04.2001)

(74) Agent: GÖSSMANN, Klemens; Internationaal Oc-  
trooibureau B.V., Prof. Holstlaan 6, NL-5656 AA  
Eindhoven (NL).

(25) Filing Language: English

(81) Designated States (national): CN, JP, KR.

(26) Publication Language: English

(30) Priority Data:  
60/201,488 3 May 2000 (03.05.2000) US  
09/621,522 21 July 2000 (21.07.2000) US(84) Designated States (regional): European patent (AT, BE,  
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,  
NL, PT, SE, TR).(71) Applicant: KONINKLIJKE PHILIPS ELECTRON-  
ICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA  
Eindhoven (NL).

## Published:

- with international search report
- before the expiration of the time limits for amending the  
claims and to be republished in the event of receipt of  
amendments

(72) Inventors: SWILLENS, Peter, J., L., A.; Prof. Holst-  
laan 6, NL-5656 AA Eindhoven (NL). MIDDELJANS,  
Jakobus; Prof. Holstlaan 6, NL-5656 AA Eindhoven  
(NL). ALBERDA, Okke; Prof. Holstlaan 6, NL-5656 AAFor two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 01/84539 A1

(54) Title: VOICE COMMANDS DEPEND ON SEMANTICS OF CONTENT INFORMATION

(57) Abstract: Voice control of the play-out or other processing of video or audio content information uses voice commands that  
semantically relate to the content information.

Voice commands depend on semantics of content information

The invention relates to voice control, especially for the play-out of content information by consumer electronics (CE) equipment.

5           Voice-controlled equipment is known from, e.g., U.S. patent 4,506,377; U.S. patent 4,558,459; U.S. patent 4,856,072; U.S. patent 5,255,326, and U.S. patent 5,950,166 all incorporated herein by reference. U.S. patent 5,255,326 in particular addresses an interactive audio system that employs a sound signal processor coupled with a microprocessor as an interactive audio control system. A pair of transceivers, operated as stereophonic  
10   loudspeakers and also as receiving microphones, are coupled with the signal processor for receiving voice commands from a principal user. The voice commands are processed to operate a variety of different devices, such as television, tape, radio or CD player for supplying signals to the processor, from which signals then are supplied to the loudspeakers of the transceivers to produce the desired sound. Additional infrared sensors may be utilized  
15   to constantly triangulate the position of the principal listener to supply signals back through the transceiver system to the processor for constantly adjusting the balance of the sound to maintain the "sweet spot" of the sound focused on the principal listener. Additional devices also may be controlled by the signal processor in response to voice commands which are matched with stored commands to produce an output from the signal processor to operate  
20   these other devices in accordance with the spoken voice commands. The system is capable of responding to voice commands simultaneously with the reproduction of stereophonic sound from any one of the sources of sound which are operated by the system.

          Speech recognition is a technology, aspects of which are discussed in, e.g., U.S. patent 5,987,409; U.S. patent 5,946,655; U.S. patent 5,613,034; U.S. patent 5,228,110;  
25   and U.S. patent 5,995,930, all incorporated herein by reference.

The known speech control and voice control of devices or applications is limited to a fixed set of commands that is tied to the equipment. The inventors have realized

that user-friendliness of, and ergonomic aspects during operational use of, voice-controllable equipment are enhanced if the voice command or voice commands are linked to the information content to be played out, rather than to the apparatus or platform. That is, the inventors believe that control of CE equipment should be content-centric, rather than device-

5 centric.

Accordingly, in one aspect of the invention, it is proposed to integrate speech commands with the content information in or on a data carrier such as a CD, a DVD or a solid state memory. The commands are preferably tailored to the semantics of the content information. For example, if the content information comprises audio, e.g., a collection of songs, selection of one or more specific ones of the songs is achieved by speaking the title or

10 part of the lyrics of the song. Special meta-data is added to the content of the CD to enable this feature. This meta-data is typically, but not necessarily, a representation of the vocabulary required by the voice controller of the device or application to enable voice control for that particular CD and the music on it. Alternatively or supplementarily, the user

15 can hum or (attempt to) sing a part of the desired piece of music in order to select it for play out. Within this context, see U.S. patent 5,963,957 issued 10/5/99 to Mark Hoffberg for BIBLIOGRAPHIC MUSIC DATA BASE WITH NORMALIZED MUSICAL THEMES (attorney docket PHA 23,241), incorporated herein by reference. This latter patent relates to an information processing system that comprises a music database. The music database stores

20 homophonic reference sequences of music notes. The reference sequences are all normalized to the same scale degree so that they can be stored lexicographically. Upon finding a match between a string of input music notes and a particular reference sequence through an N-ary query, the system provides bibliographic information associated with the matching reference sequence. This system can also be used to convert the input hummed by the user into a play

25 command via the N-ary query.

Without further measures the audio output of the system may trigger an undesirable activation of the speech-controlled processing, e.g., when a song is being played out. This undesirable activation is prevented, e.g., through echo cancellation, by pressing an activation button on the remote, e.g., the Pronto (TM), the universal programmable remote

30 from Philips Electronics, to activate speech command receipt, or by having the equipment registering the user making a specific gesture, etc. If the content information comprises video, key scenes are labeled by key words so that speaking those words sets the playing out at the start of the relevant scene. A key word profile of the video content may be used to identify certain scenes, either through a one-to-one mapping of the user's voice input to the

keywords or through a semantic mapping of the user's voice input onto an indexed list of the content's keyword labels and their synonyms. Preferably, undesired activation is prevented from occurring, e.g., by using certain fixed commands or parts thereof such as a prefix. Similarly, interactive software applications using graphics, e.g., virtual reality or video games, are made speech-controllable by allowing the processes to associate speech input with controllable features of graphics objects displayed or to be displayed. For example, actions to be carried out by a graphics object, e.g., an avatar, are made speech-controllable or speech-selectable by having the user say the proper words fitting the semantic context. This is suitable for video games allowing multiple modalities of control (e.g., both hand-input through joy-stick and speech input), as well as educational programs for teaching another language, or for teaching children the proper words and expressions for certain concepts such as tangible objects or actions. The speech is converted into data for being processed so as to identify the proper action intended. This is achieved through, e.g., semantic matching of the speech data with items in a pre-determined look-up table and finding the candidate for the closest match. The association between speech input and action intended may be made trainable by virtue of taking user-history into account.

In another aspect of the invention, speech commands are derived from the content when the content is stored locally after downloading from the Web and/or playing-out. For example, key words in the lyrics are identified and stored as associated with the piece of audio whereto they pertain. This can be done by a dedicated software application. Either the digital data are analyzed or the audible lyrics are analyzed during the first play out of the audio content, for example, by isolating the voice part from the instrumental part and analyzing the former. The speech commands thus created can be used in addition to, or instead of, the basic set that comes with the specific content.

In yet another aspect of the invention, the user is enabled to download pre-existing or customized commands from the Web that pertain to specific content information and that are to be stored at the user's equipment as semantically associated with the information content for the purpose of enabling voice control. Thus, the user can make his/her home library of electronic content information, considered as a resource for the home network, fully speech driven. For example, the user has a collection of CD's, DVD's, in his/her jukebox and/or on a hard disk. If the content relates to publicly available audio and video, a service provider can create a library of annotations for each piece of the content in advance, and the user can download those elements that are relevant to his/her collection. The annotations for a CD or DVD can be tied to the disk's identifier as well as to its segments.

For example, the name of an album, spoken by the user, is linked to a certain identifier that in turn enables retrieval and selection of the CD or DVD in the jukebox. The name of a song or scene can be linked to both the identifier of the CD or DVD and to the relevant key frames.

The user then speaks the terms "movie" and "car chase" and gets in return the movies

5 available that have scenes in them that relate to a car chase.

In yet another aspect of the invention, the speech commands are linked to the content as presented in an electronic program guide (EPG), e.g., as broadcast by a service provider. Again, a speech interface enables to select a specific program or program category that matches or match the words spoken by the user.

10 In yet another aspect of the invention, commands as spoken by the user are processed via a server, e.g., a home server or a server on the Web and routed back to the Web-enabled play-out equipment as instructions. The server has an inventory of content available and a dictionary of words that are representative of the content's semantics. The Web-enabled equipment identifies to the server the content, e.g., through the identifier code  
15 of a CD or DVD, or through the header of a file, whereupon the speech commands for this content are readily matched to instructions for the control through, e.g., a look-up table.

The voice control enables, e.g., the selection of a piece of content information for play-out, or for storage or for fast forward until a stop, etc. Also, content bookmarked with key words in advance can be browsed under voice control for retrieval of certain  
20 excerpts matching the voice input at the key word level.

Another aspect of the invention addresses copying the content information from one storage medium, e.g., a CD or DVD, onto another storage medium. The first storage medium comprises the content information and the control information that enables voice control as explained above. Preferably, the information for the voice control is copy-  
25 protected, as a result of which the copy does not have the control commands. This is considered a feature supporting the content information industry. If the consumer wants to have a full copy of the voice controlled version, he or she can download the voice control information from a server on the Internet identified by a link to the CD number or DVD number, at a certain price. This has the advantage that the author's rights are acknowledged,  
30 even if the price is merely symbolic. Thus, this feature contributes to maintaining awareness that content information is the intellectual property of the author or his/her assignees.

Incorporated by reference herein is U.S. serial no.09/345,339 (attorney docket PHA 23,700) filed 7/1/99 for Mark Hoffberg and Eugene Shteyn for CONTENT-DRIVEN SPEECH- OR AUDIO-BROWSER. This patent document relates to searching the Internet in

order to find resources that provide streamable audio such as live Internet broadcasts. The resources are identified based on their file extension and are categorized according to, e.g., the natural language or music style. The user is enabled to browse the collection based on textual or musical input.

5                   The expression "voice command" as used herein is meant to indicate a voice control input that may consist of one or more keywords but it may also comprise a more verbose linguistic expression.

                  The invention is in further detail, and by way of example, with reference to the accompanying drawing, wherein:

10

Figs.1 and 2 are block diagrams of systems in the invention.

15                   The invention allows for voice control of apparatus or software applications, in particular of those that use content pre-recorded on a storage medium. Voice commands are used that semantically relate to, are associated with or based on, the content as stored in the storage medium. The commands are therefore different per sample of the medium's content. For example, the commands available for a CD with music from composer or lyrics  
20 author X are different from those for a CD with music composed by composer or lyrics author Y.

                  For a CD player, the operation is as follows. The user inserts a CD of performer Daan van Schooneveld into the player. The CD stores the music and the software to enable the user to interact with the CD through voice control. When the user says  
25 "Mustang Danny", the player starts to play the rock song of that title, one of the tracks of Schooneveld's CD. When the user says "leaking oil", the player starts playing the blues song whose lyrics has the line "I wept gently in the rain as the gearbox was still leaking oil". And so on. A similar control scenario applies to the voice control of a set top box or another apparatus that has a CD drive. A user-programmable delay may be needed between voice  
30 commands to separate the commands per song. Alternatively, specific expressions can be used to serve as a divider between commands per song. For example, the user may say: "Mustang Danny play twice, Leaking oil play once; ". This gets interpreted as that the song "Mustang Danny" is to be played out twice in succession, then the song relating to the "leaking oil" is to be played twice in succession. The expressions "play twice" and "play

once" serve as dividers to identify each song and what the system is supposed to do with it before the system prepares for receipt of another voice command.

Voice control of a jukebox application on a PC is illustrated as follows. A jukebox application is a software application that allows for archiving CD content on the PC's hard disk drive (HDD). The user has archived the Jos Swillens "Greatest Hits" CD on the HDD. When the user says "Swil, Beemer", the jukebox starts to play "My Beemer fits my crewcut", one of the tracks of Swillens' CD archived on the PC. The voice commands need not consist of only keywords but may comprise more verbose linguistic expressions. For example, the user may say "play from Swillens' greatest hits the title about the crewcut", the system processes the voice input to match it with one of the options available using, e.g., a suitable search algorithm in an index list. When the user says "Swil, always be nice to your patent attorney", the jukebox starts playing the symphonic classic "Always be nice etc."

The user has also archived the "Greatest Hits" CD from Koos Middeljans on the PC. When the user says "Koos, Sweet Dommel Valley ", the jukebox starts to play the folk song with that title, one of the tracks of the CD archived. When the user says "Koos, Nat the Lab", another track of Mid's "Greatest Hits" CD archived on the PC, the jukebox starts playing "Nat the Lab". When the user says "Middeljans, greatest hits, random", the jukebox starts playing the tracks of this CD in a random order.

Content protection in terms of copyright is a sensitive issue. Copy protection measures are available and implemented, e.g., DRM (Digital Rights Management). To contribute to this, the speech commands as supplied together with the semantically related content information on a CD or DVD could be implemented in such a manner that they cannot be copied to a location other than the onboard memory of a player. Any copy to another location would lose this feature and become less attractive.

In another example, the user downloads the content via the Internet together with the semantically related control data that enables voice controlled selection and play out in a similar manner as discussed for the jukebox. The control data is preferably an integral part of the downloaded data in this example.

For background on jukebox technology, see U.S. serial no. 09/326,506 (attorney docket PHA 23,417) filed 6/4/99 for Pieter van der Meulen for VIRTUAL JUKEBOX, herein incorporated by reference.

The same content information can be tied to phonetically different sets of voice commands, for example, to allow for differences in language and in pronunciation in different geographic regions so as to facilitate voice recognition. Within this context, the user

preferably has a choice of the language he or she wants to use for voice control of the system. The storage medium may have too small a storage capacity for storing the commands of all the languages likely to be used. If voice commands are not available from the medium in one of the languages most likely to be used, the play out device is preferably able to download the equivalent speech commands in the desired language whereupon the system will translate the commands at run time into the corresponding instructions. A dedicated service can be made available on the Internet. Within this context, reference is made to U.S. serial no. 09/160,490 (attorney docket PHA 23,500) filed 9/25/98 for Adrian Turner et al., for CUSTOMIZED UPGRADING OF INTERNET-ENABLED DEVICES BASED ON USER-PROFILE (SmartConnect (TM)), and to U.S. serial no. 09/519,546 (attorney docket US000014) filed 3/6/00 for Erik Ekkel et al., for PERSONALIZING CE EQUIPMENT CONFIGURATION AT SERVER VIA WEB-ENABLED DEVICE, both incorporated herein by reference. These documents discuss services provided to CE end-users via the Internet.

It is expected that in the future audio and video content will be supplied to the end-user via the Internet to an ever larger extent. The recording is then accomplished at home under secure circumstances. The local recording preferably allows the consumer to create his/her own command set semantically related to a specific piece of content information. This needs some editing and a preferably a specific graphical user interface (GUI) that assists the user with establishing the relationships between content segments, voice input commands and actions or processing desired. For example, if the content information is not annotated at all, the user has to specify which segments he/she wants to control as separate items, how he/she wants to control is with what voice commands, and what actions should be taken upon what segment under what command. Once created, the command set can be stored together with the specific content in the same file or linked with the specific content using a unique identifier.

In a more sophisticated system, the phonetic transcription covers any relevant form of phonetic transcription, independent of phoneme inventory, for example, limited to a subset of the vocabulary, or just for the exception of a standard pronunciation. Mutatis mutandis, this also applies to an optional acoustic model (acoustic references). A language model can be used optionally, that includes a description of how people typically interact with the system and say sentences (the so-called "language model"), be it via example sentences, patterns or phrases, via (stochastic) finite state grammars, via (stochastic) context free grammars, or another kind of grammar. The language model may just contain a modification of any standard way of communicating. As to speech understanding, the system



optionally includes any description of what action should be triggered by certain words, commands, phrases, expressions, typically as given via a grammar. The system may include a dialogue model that includes a description of how the system should react to user's input and how the system enters a dialogue mode. For example, the system may ask for clarification, or  
5 to reconfirm a command, etc., under specific circumstances. The system may use a relationship between the data configuring the speech recognizer and other data. For example, the system has a display that shows what the user can say in order to play a current track.

Preferably, the storage medium, e.g., a CD, DVD, solid state (e.g., flash) memory, etc., has a bit pattern that gets recognized during start-up and that confirms the  
10 availability of the voice command feature. The confirmation can be conveyed to the user through, e.g., a pop-up screen on a display or spoken pre-recorded text supplied via the loudspeakers.

As to the formatting of the voice control software in the medium, CD-DA has the extra capacity of the R - W channels that can be used for adding the voice command  
15 feature without losing the CD's backwards compatibility. The lead-in tracks may not have adequate storage for the various language versions, but the data can be downloaded from the disc into a local memory. In this case each language has to be only once on the disc. CD ROM, on the other hand, has a file structure which makes it easy to accommodate the speech control file on the disc as required. DVD also has a file structure and allows for the same  
20 approach as the CD ROM. Flash, HDD etc can be handled in the same way.

Fig.1 is a block diagram of a system 100 in the invention. System 100 comprises a play-out apparatus 102 for playing out content information 104 stored on a carrier 106. Carrier 106 comprises, for example, a CD, a DVD or a solid state memory. Alternatively, carrier 106 comprises a HDD onto which content information 104 has been  
25 downloaded via the Internet or another data network. Content information 104 in these examples is stored in a digital format. As is clear to the person skilled in the art, content information 104 may also be stored in an analog format. Apparatus 102 has a rendering sub-system 108 for making content information 104 available to the end-user. For example, if content information 104 comprises audio, sub-system 108 comprises one or more  
30 loudspeakers, and in case content information 104 comprises video information sub-system 108 comprises a display monitor.

According to the invention, carrier 106 comprises control information 110 that is semantically associated with content information 104. Control information 110 enables a data processing sub-system 112 to determine if a voice input 114 by the user via a

microphone (not shown) matches an information item in the control information. If there is a match, the relevant play-out mode is selected, examples of which have been given above. The semantic relationship between control information 110 on the one hand, and content information 104 on the other hand facilitates user-interaction with apparatus 102, owing to  
5 the highly intuitive correspondence, as explained above in the play-out examples of audio content. Preferably, visual feedback is provided via a local display, e.g., a small LCD 116, as to the content available and/or mode selected.

Carrier 106 can be a component that can be inserted into apparatus 102 one at a time. Alternatively, apparatus 102 comprises a jukebox functionality 118 that enables to  
10 select content from among multiple carriers (not shown) like carrier 106 or from among even physically different ones, CD and solid state memory, for example.

Control information 110 is shown here as stored or recorded with content information 104 on carrier 106. A CD, DVD or flash can thus be supplied having pre-recorded voice control applications and commands. Alternatively, control information 110  
15 cooperates with a dedicated software application running on data processing system 112 for matching voice input 114 with one or more items available in control information 110. In this latter configuration, the software application is provided via another channel than the control information, e.g., via the Internet or a set-up diskette for setting up apparatus 102.

Voice control itself is known, and so is user-interaction with an apparatus for  
20 selecting an operational mode of the apparatus. The invention here relates to using a control interface, part of which is semantically associated with the content information available for playing-out.

Options that are preferably integrated within a system of the invention include the following. System 100 provides auditory or visual feedback in response to the user  
25 having entered a spoken command. For example, system 100 confirms receipt of the command, e.g., by repeating the command word or command words in a pre-recorded voice if there is a match, or by supplying the word "confirmed" in a pre-recorded voice if there is a match. This feature can be readily implemented with a relatively small number of predetermined commands per information content item. The confirmation data can be  
30 integrated within control data 110. If the voice command as given by the user is not understood, i.e., system 100 does not recognize this and does not find a match in control data 110, system 100 supplies auditory feedback indicating the negative status. For example, system 100 supplies in a pre-recorded voice "cannot process this command", "cannot find this artist", or cannot find this song" or words of a similar meaning. Instead of, or in addition

to, auditory feedback, system 100 can give visual feedback, e.g., a green blinking light if system 100 is capable of processing the voice input, and a red light if it is not. Along the same lines, system 100 preferably pronounces, in a pre-recorded or synthetic voice, the name of the artist and the song title or album title of the content selected for being played out. The synthetic voice uses a text-to-speech engine for this feature so the system can use the information that comes available from the download or the media carrier. Text-to-Speech (TTS) systems convert words from a computer document (e.g., a word processor document, a web page) into audible speech through a loudspeaker. In a TTS system, preferably the words are stored together with their phonetic transcription, comprising intonation of carrier sentences, etc. Also, as an option, control data 110 comprises pre-recorded or synthetic voice data explaining to the user which commands, e.g., which song keywords, are available. The pre-recorded or synthetic voice data can again be part of control data 110. The user should be able to turn this on or off when he/she does not want the system to provide auditory feedback.

Fig.2 is a diagram illustrating a system 200 with an EPG wherein available content information is identified and arranged in rows 202 and columns 204 on a display monitor 206. For example, each respective row represents a respective TV channel and each of the columns represents a specific time slot. At the intersection of each specific row and column pair, e.g., row 208 and column 210, a label or title 212 is shown that represents the content available from that specific channel and in that particular time slot. Other types of arrangements can be used instead, e.g., by topical category and time, or ranked by user-preference according to a profile per channel or resource (e.g., on the Internet), etc. The user can browse the EPG by, e.g., moving a window 214 across the grid of the EPG through a suitable user-interface (e.g., arrow keys on a wireless keyboard or another directional device, not shown) in order to get the portion of the EPG displayed that falls within the boundaries of window 214. The user can thereupon select particular content information by clicking or highlighting the associated label in the portion displayed.

Typically, an EPG is supplied via the Internet by a service provider. In the invention, the EPG is enhanced with additional control software 216 that enables a mode of user-interaction with the EPG other than the conventional clicking or highlighting of a desired label. Control software 216 is preferably downloaded, updated or refreshed together with the EPG. Control software 216 comprises control information 218 associated with the semantics of the labels that identify the programs in the EPG for user-selection. For example, when the user inputs the expression "movies" into data processing sub-system through user-input device 220, e.g., by voice input through a microphone, the EPG's grid is re-organized

to only show the available programs according to the category "movie" in window 214, or the movie programs are graphically represented as distinct from programs in the other categories. The user then browses through the category "movies", preferably also under speech command. The user sees the movie of his/her liking and enters as voice input the expression "The Magnificent Six and Okke", the title indicated in the EPG of the classic movie about an aviation event. In another example, the user enters "tonight" and "from eight o'clock" upon which window 214 is being located to, at least partly, show the collection of programs available that day and as from eight o'clock (8:00pm) on. In yet another example, the user has identified an interesting program in the portion of the EPG displayed in window 214 and speaks the words, representative of the title of the program, into microphone 220. Then, the user speaks "watch" or "record". The words that represent the title are converted into a suitable format for comparison with control information 218. Upon finding a match, the control software 216 enables a microprocessor 222 to control a tuner 224 and display monitor 206 or a recording device 226. In this manner, the user can interact with the EPG using voice control.

## CLAIMS:

1. A method of enabling an end-user to control processing of content information, the method comprising processing a speech command that is semantically associated with the content information to be processed.
- 5 2. The method of claim 1, comprising supplying speech control software together with the information content.
3. The method of claim 1, wherein the command identifies the content information for processing.
- 10 4. The method of claim 1, wherein the content information comprises audio; and the command comprises a word occurring in the audio.
5. The method of claim 1, wherein the content information comprises video  
15 information; and the command identifies an event or object in the video.
6. The method of claim 1, wherein the content information is stored in a storage medium; and the command is stored in the storage medium for control of the processing.
- 20 7. The method of claim 1, comprising supplying feedback to the end-user regarding a status of the processing of the speech command.
8. A storage medium with content information and with data representative of a speech command for enabling an end-user to control processing of the content information  
25 through speech.
9. The medium of claim 8, wherein the speech command is semantically related to the content information.

10. The medium of claim 8, comprising at least one of the following: an optical disk; a magnetic disk; a solid state memory.
11. An electronic apparatus for processing content information, the apparatus  
5 comprising:
- a speech input for receipt of a speech command;
  - an input for receipt of a storage medium that comprises the content information and control software specific to semantics of the content information; and
  - a data processor for the processing of the content information via the software under  
10 control of the speech command.
12. The apparatus of claim 11, wherein the data processor processes the content information in response to a speech command semantically related to the content information.
- 15 13. The apparatus of claim 11, wherein the storage medium comprises at least one of the following: an optical disk; a magnetic disk; a solid state memory.
14. The apparatus of claim 11, comprising an output for indicating to an end-user  
20 a status of a processing of the voice command.
15. A method of supplying control data associated with semantics of specific content information for enabling an end-user to control processing of the specific content information through speech control as supported by the control data.
- 25 16. The method of claim 15, comprising enabling a user to download the control data via a data network.
17. The method of claim 15, wherein the downloaded control data is for use with a  
30 copy of the specific content information.
18. The method of claim 15, comprising enabling the user to download the content information via a data network.

19. The method of claim 15, wherein the content information comprises an EPG, and wherein the processing comprises interacting with the EPG.

20. An EPG comprising control data specific to semantics of content information  
5 represented by a program listing and operative to enable an end-user to interact with the EPG using speech input.

21. The EPG of claim 20 comprising software for control of supplying feedback to the end-user regarding a status of a processing of the speech input.

10

22. For an EPG, control data specific to semantics of content information represented by a program listing and operative to enable an end-user to interact with the EPG using speech input.

15 23. Speech command for control of electronic processing content information, the command being determined by semantics of the content information.

1/2

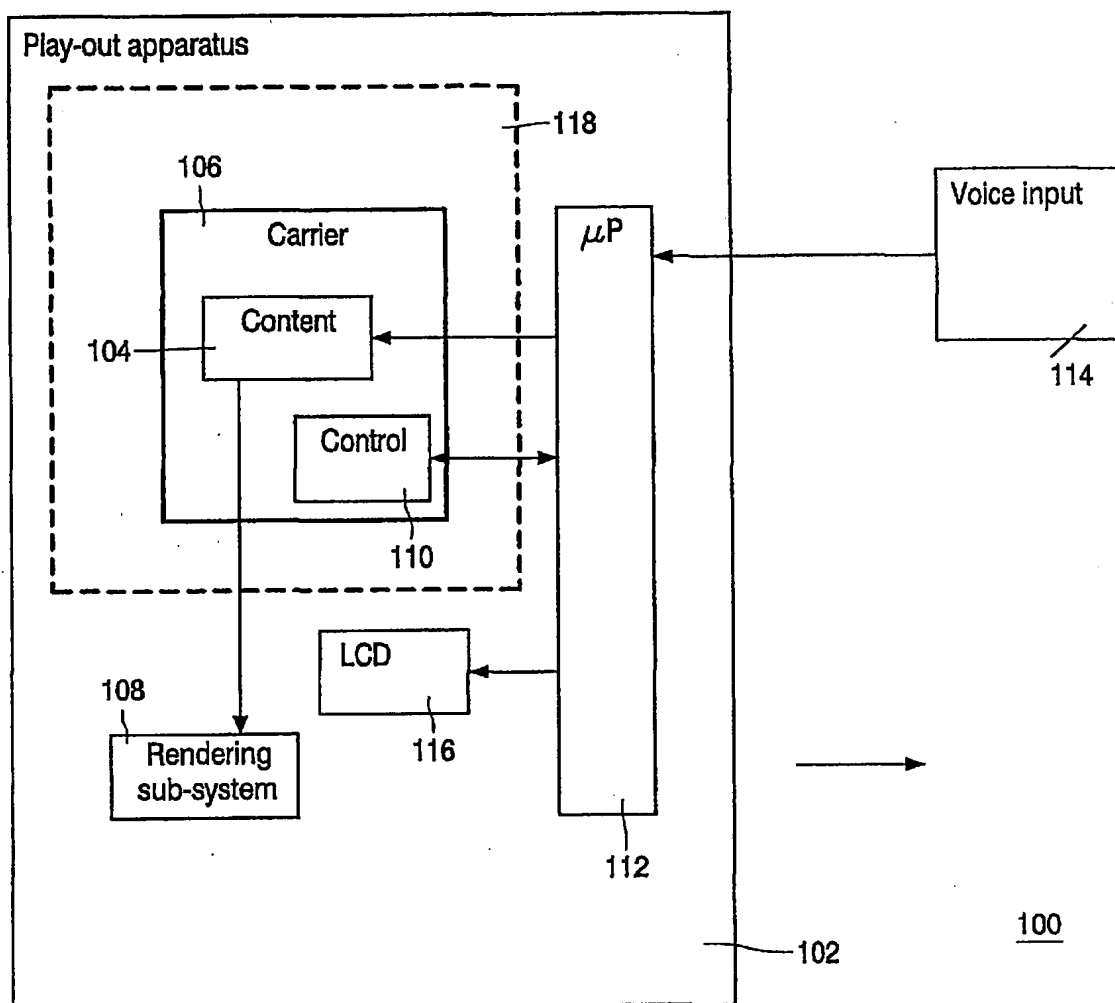


FIG. 1



2/2

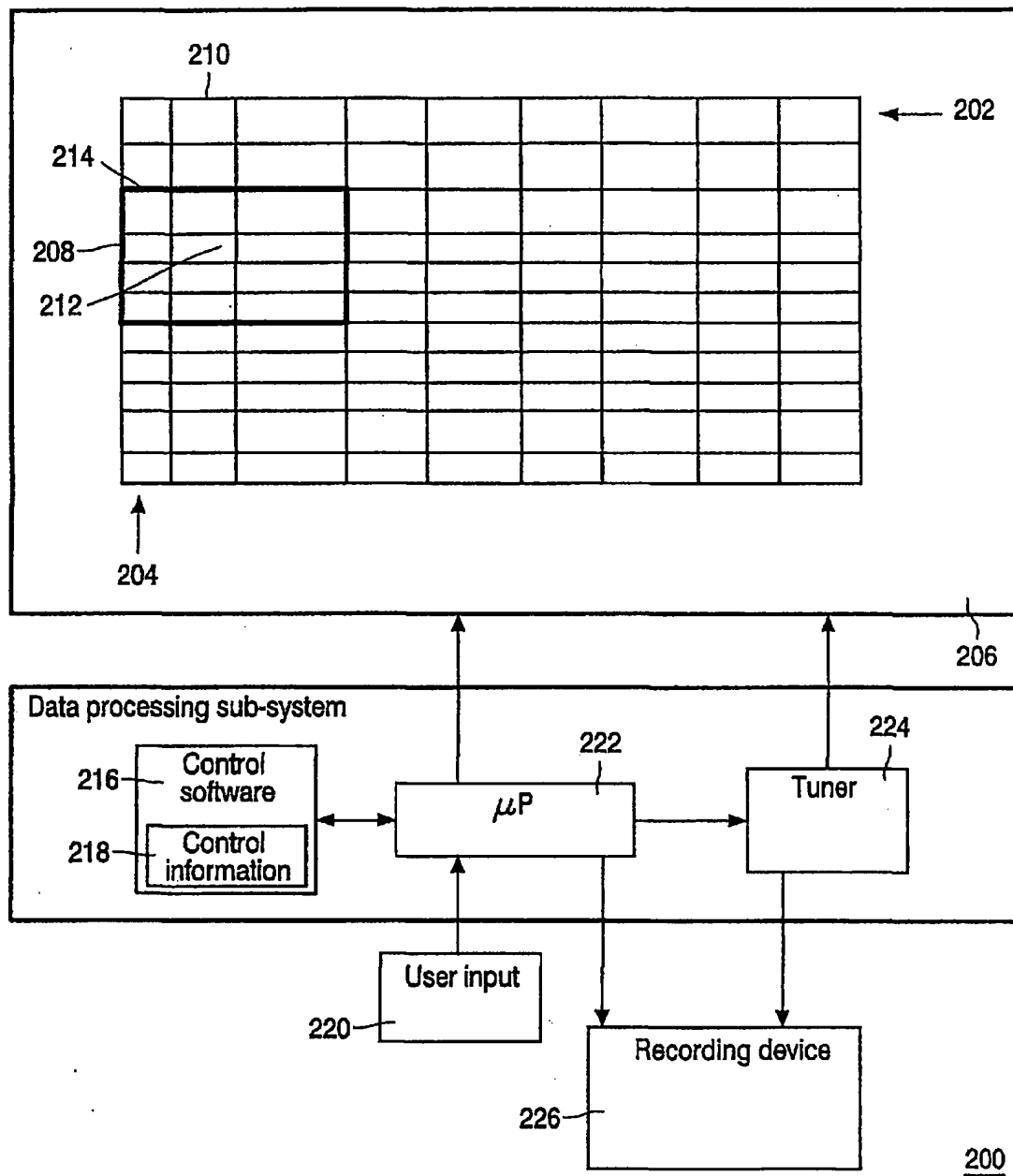


FIG. 2

# INTERNATIONAL SEARCH REPORT

national Application No

PCT/EP 01/04714

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G10L15/26

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 774 859 A (HOUSER PETER B ET AL) 30 June 1998 (1998-06-30) column 5, line 39 -column 6, line 14; figure 1 column 7, line 25 - line 33 column 29, line 23 -column 31, line 19	1-23
X,P	EP 1 037 463 A (MATSUSHITA ELECTRIC IND CO LTD) 20 September 2000 (2000-09-20) abstract page 2, column 2, line 47 -page 3, column 4, line 5 page 3, column 4, line 52 -page 5, column 7, line 8 --- -/--	1-23

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*A\* document member of the same patent family

Date of the actual completion of the international search

8. October 2001

Date of mailing of the international search report

16/10/2001

Name and mailing address of the ISA  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Wanzeele, R

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP 01/04714

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	EP 1 079 371 A (MATSUSHITA ELECTRIC IND CO LTD) 28 February 2001 (2001-02-28) abstract page 2, column 2, line 32 - line 56 page 3, column 3, line 27 - line 56 page 3, column 4, line 42 - line 53	1-23

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 01/04714

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
US 5774859	A	30-06-1998	AU	4748896 A	31-07-1996
			WO	9621990 A2	18-07-1996
EP 1037463	A	20-09-2000	EP	1037463 A2	20-09-2000
EP 1079371	A	28-02-2001	EP	1079371 A1	28-02-2001
			JP	2001148814 A	29-05-2001